DOCUMENT RESUME

ED 282 913 TM 870 326

AUTHOR Naron, Nancy K.; Elliot, Norbert
TITLE Using Young Children's Writing Samples in Program Evaluation.
INSTITUTION Fort Worth Independent School District, Tex. Dept. of Research and Evaluation.
PUB DATE Apr 87
NOTE 73p.; Paper presented at the Annual Meeting of the American Educational Research Association (Washington, DC, April 20-24, 1987). Appendices D and E may not reproduce well.
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Comparative Testing; *Computer Assisted Instruction; Computer Software; Conventional Instruction; Grade 1; Holistic Evaluation; Kindergarten Children; Primary Education; *Program Evaluation; Reading Instruction; School Districts; *Writing Evaluation; Writing Instruction; *Writing Processes; *Writing Skills
IDENTIFIERS *Childrens Writing; Fort Worth Independent School District TX; Writing Samples; *Writing to Read

ABSTRACT

A writing assessment study was conducted in the Fort Worth Independent School District (Texas) in the spring of 1986 as part of a larger program evaluation study of the Writing to Read (WTR) Program. WTR is a computer-based instructional system designed to develop the writing and reading skills of kindergarten and first-grade students. Two types of comparisons were conducted to address the effectiveness of WTR: (1) writing skills of WTR participants with those of students of the same grade level in traditional classrooms; and (2) the writing skills of WTR participants with those of students whose teachers had been trained in Writing Process (WP) instruction. A total of 215 kindergarten and 270 first-grade writing samples were collected. The samples were scored by kindergarten and first-grade teachers. The results indicated no differences between the writing samples of first-grade students in WTR and traditional classes, but first-graders in WP classes scored significantly higher than those in the other two groups. When the scores of kindergarten students in WTR, WP, and traditional classrooms were compared, the WTR students scored highest, WP students next highest, and traditional students scored the lowest. However, the sample selection for this comparison carried some gross limitations. The comparison between writing samples of WTR and traditional kindergartners, which did not have sample selection constraints, indicated no significant differences between the writing scores of these two groups. Appendices include: (1) instructions for collecting writing samples; (2) scoring criteria; and (3) examples of writing samples. (Author/JAZ)

FORT WORTH INDEPENDENT SCHOOL DISTRICT
FORT WORTH, TEXAS

ED282913

Using Young Children's Writing Samples in Program Evaluation

DEPARTMENT OF RESEARCH AND EVALUATION

TM 810 324

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
N. Naron
D. Powell
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

Using Young Children's Writing Samples in Program Evaluation

Nancy K. Naron
Department of Research and Evaluation
Fort Worth Independent School District

and

Norbert Elliot
Department of Literature and Languages
East Texas State University

TM 870 326

Paper presented at the annual meeting of the American Educational Research Association, Washington, D.C., April, 1987.

BEST COPY AVAILABLE

Using Young Children's Writing Samples in Program Evaluation

Nancy K. Naron and Norbert Elliot

Abstract

This paper describes a writing assessment study that was conducted in the spring of 1986 as part of a larger program evaluation study of the Writing to Read (WTR) Program. WTR is a computer - based instructional system designed to develop the writing and reading skills of kindergarten and first-grade students. It was implemented in 16 schools in Fort Worth, Texas, in the fall of 1985.

Two types of comparisons were conducted to address the effectiveness of WTR in the area of writing skills. The first compared the writing skills of WTR participants with those of students of the same grade level in traditional classrooms. The second comparison was between the writing skills of WTR participants and those of students whose teachers had been trained in Writing Process (WP) instruction.

A total of 215 kindergarten and 270 first-grade writing samples were collected. The samples were scored by kindergarten and first-grade teachers who received training just prior to the scoring session. A somewhat holistic scoring approach was used, one which focused on the content and quality of the writing rather than the surface features, but allowed the readers time necessary to decipher the phonetic spelling and other irregularities inherent in young children's writing.

The results indicated no differences between the writing samples of first-grade students in WTR and traditional classes, but first-graders in WP classes scored significantly higher than those in the other two groups. When the scores of kindergarten students in WTR, WP, and traditional classrooms were compared, the WTR students scored highest, WP students next highest, and traditional students scored the lowest. However, the sample selection for this comparison carried some gross limitations. The comparison between writing samples of WTR and traditional kindergartners, which did not have sample selection constraints, indicated no significant differences between the writing scores of these two groups.

The methods and results of this study are compared to those of similar studies. In addition, the paper discusses aspects to be considered when using young children's writing samples in program evaluation.

TABLE OF CONTENTS

Introduction 1

Description of the Writing to Read Program 2

Description of Writing Process Instruction 3

Comparison of WTR and WP 4

Development of Writing Assessment Techniques 5

Concerns in Scoring Young Children's Writing 5

Other Evaluations of Writing to Read 6

Implementation of WTR in Fort Worth Schools. 7
- Description of the WTR Schools 7
- Cost of the Program 7

Implementation of Writing Process in FWISD 8
- Description of the WP Classes 8
- Cost 8

The FWISD Writing Assessment Study. 8
- Design of the Study 8
- Collection of Writing Samples 11
- Scoring of the Writing Samples 13
- Results of the Writing Assessment 16
- Discussion of the Writing Assessment Results 20

Conclusion. 22

References 24

Tables 25

Appendices. 37

Using Young Children's Writing Samples in Program Evaluation

Nancy K. Naron and Norbert Elliot

INTRODUCTION

In the fall of 1985, the Fort Worth Independent School District (FWISD) implemented the use of the Writing to Read (WTR) Program in 16 elementary schools. This program, which is marketed by the International Business Machines Corporation (IBM), is a computer-based instructional system designed to develop the writing and reading skills of kindergarten and first-grade students. One of the main goals of WTR is to help the students develop their ability to express their ideas on paper. The objective of this study was to determine the effectiveness of the WTR Program in the development of students' writing skills, in comparison to students of similar socioeconomic status in other types of classrooms.

Previous evaluations of WTR (Guttinger, 1986; Murphy and Appel, 1984) have all compared the effects of WTR with those of traditional instruction, and have found differences between the writing skills of the two groups. As pointed out by Norton (1984), there is always a question about what a new educational program should be compared with when it is implemented. It seems inappropriate to compare it to the complete absence of a program, and yet, writing has not traditionally been emphasized in the early elementary years. Norton admitted that "...some doubt exists concerning what the Writing to Read program was actually being compared with" (p.23), and this doubt applies to other evaluations of the WTR program as well.

Therefore, a decision was made in the present study to compare the effects of WTR not only to those of traditional classes, but also to those of classes known to be teaching writing in a strong and consistent fashion. FWISD is

fortunate to be one of the sites sponsored by the National Writing Project. As part of this project, teachers receive a minimum of 30 hours of training in Writing Process (WP) instruction. Students of teachers trained in WP were selected to be a second comparison group.

DESCRIPTION OF THE WRITING TO READ PROGRAM

Writing to Read (WTR) is a computer-based instructional system designed to develop the writing and reading skills of kindergarten and first-grade students. The objectives of the program are to help the students develop their ability to express their ideas on paper and to read what they have written.

The program is provided within the context of a planned learning center, which is staffed by a full-time aide. The students spend from 30 to 60 minutes per day, accompanied by their teacher, using the equipment and materials in the WTR center. This center, or lab, is organized into at least five learning stations according to guidelines outlined by the WTR developers.

One of the main learning stations in the WTR lab is the Computer Station. The aide directs the students' activities at this station. Students proceed through a series of ten instructional cycles which teach some basic vocabulary words using a phonemic spelling system. In addition, practice activities are available on the computer to reinforce the skills learned in the instructional cycles. A great deal of repetition is a central part of the computer activities.

A second learning station is the Work Journal Station. The Work Journals are designed to provide additional support and practice in the learning of the material in the ten instructional cycles. Some of the pages are designed to be completed in conjunction with an audio tape. The main activity is to practice writing the cycle words in a number of formats.

A third work station is the Writing/Typing Station. In one area of this station, students write stories by hand. In the other area of this station,

students type their stories on computer using a word processing program.

In a fourth work station, called the Listening Library Station, the students have the opportunity to match speech with written language. They listen to stories recorded at a slow pace on tape while following the written text in the corresponding book.

The fifth required station is the Make Words Station. Various activities are arranged to have the children practice matching letters and sounds using manipulative and appropriate alphabet materials.

DESCRIPTION OF WRITING PROCESS INSTRUCTION

Writing Process is a system of writing instruction which addresses the process by which students generate content. Formerly, writing was viewed as product-oriented. That is, instruction in writing typically took place after the writing was complete. A great deal of research that has been generated over the past two decades (Emig, 1977; Graves, 1981), however, has shown that writing instruction conducted in that traditional fashion does not improve student writing.

The research has indicated that for student writing to improve, no matter the grade level, instruction must occur throughout the series of discrete stages in which writing develops: the initial drafting of a piece of discourse, peer response to first draft, revision, redrafting, editing, conferencing, the final draft, publishing, and evaluation. These activities can generally be categorized as prewriting, writing, sharing, revising, editing, and evaluating. The process is highly individualistic and recursive in nature. With the very young writer, less emphasis is placed on the revising, editing, and evaluating stages than on the other stages.

The student/writer depends heavily on peer response to shape meaning and to address correctness. The teacher's role is that of collaborator/editor. In this

way, writing process is student centered, rather than teacher centered, and a great deal of interaction occurs among the students and between student and teacher throughout the writing process. As a result, the children are participating in fundamental learning strategies which provide reinforcement, feedback, connection, and personalization (Emig, 1977).

Writing process instruction is the model being utilized and constantly expanded and improved by the National Writing Project (NWP). The first NWP site was at Berkeley, known as the Bay Area Writing Project. There are now 146 NWP sites nationwide using this model. FWISD is one of these sites and the only one located in a public school district.

A fundamental tenet of the National Writing Project is that teachers of writing must themselves write; therefore, teachers of Writing as Process know the sort of problems that all writers face in that they themselves are practitioners of their craft. In addition, teachers are encouraged to make presentations to other teachers and to meet regularly in order to discuss and refine their interests is the teaching of writing.

COMPARISON OF WTR AND WP

The philosophy behind WP is similar to that of WTR in that young children are allowed to spell phonetically, with a concern toward the quality of the content of the writing over the editorial aspects (e.g., spelling, punctuation). There are major differences between the programs, however. WP encourages a great deal of social interaction among the children and between teacher and student to help the children progress through the developmental stages of writing. In WTR, the social interaction aspect is minimal compared to the reliance on the computer and other materials. On the other hand, WTR includes components designed to teach reading skills, while WP focuses on the development of writing skills. In both programs, students are strongly encouraged to sound out letters when

attempting to write new words.

DEVELOPMENT OF WRITING ASSESSMENT TECHNIQUES

Early efforts toward developing empirically sound methods of the assessment of writing ability were generally either abandoned or considered much too unreliable or invalid to yield much information. However, in 1966, Godshalk and his colleagues attempted to correlate multiple choice testing of writing skills with direct assessment of writing. Ironically, it was in their efforts to validate multiple choice testing that they discovered the method of scoring that today is known as holistic, or impressionistic, scoring. The form of holistic scoring as it is presently used in massive assessment programs such as the New Jersey Basic Skills Placement Test and in individual program assessments throughout the nation was refined and disseminated by various researchers (Odell and Cooper, 1977; Myers, 1981; White, 1985).

As it is presently defined, holistic scoring is conducted through a series of steps: selecting sample papers which exhibit a range--commonly six points--of writing, reproducing the sample papers, and training readers to come to consensus on the values of these papers. Once consensus is achieved, each of the remaining writing samples is scored by two readers. Papers receiving scores exceeding the pre-established range of agreement are termed discrepant and resolved by a third reader.

CONCERNS IN SCORING YOUNG CHILDREN'S WRITING

The evaluation of young children's writing poses some rather unique problems. Since phonetic spelling is characteristic of kindergarten and first-grade students, for instance, a truly holistic reading is impossible. The writing must be "translated" by the reader. Moreover, readers must strive even harder to go beyond the surface features of correctness and understand that young

writers commonly employ irregularities in penmanship, grammar, and punctuation. Designers of the assessment must train readers to focus on the organization of ideas, for example, or extended detail achieved by the writer's fluency. Readers must accordingly come to recognize such values and score them appropriately. Hence, totally impressionistic scoring is shifted towards a more focused scoring so that the young writers may be rewarded for experimentation in their newly acquired technology.

Too, the design of the topic or prompt for the writing assessment is a unique problem. Since formal evaluation of this population is such a new concept, there are presently no studies which deal with young children's abilities to handle subtleties of audience and rhetorical modes of composition. Therefore, researchers are safest to design prompts which evoke the narrative patterns common to children's speech: those which evoke concrete, narrative experiences which are believed to be common to the cognitive capabilities of beginning writers.

OTHER EVALUATIONS OF WRITING TO READ

The first formal empirical evaluation of WTR was conducted by the Educational Testing Service (ETS) from 1982 to 1984 (Murphy and Appel, 1984), and financed by IBM. The writing assessment study that was part of their larger evaluation compared some 6,000 writing samples from WTR and non-WTR children and found that the WTR students "clearly surpassed comparison students in writing performance" (9.4). Although the ETS study followed the holistic scoring procedures and concerns for scoring young children's writing outlined above, there has been some criticism of several other aspects of the ETS study (Hathaway, 1985).

Other evaluations of WTR that have included writing assessments (Guttinger, 1986; Karam, 1986) have not utilized the rigorous procedures of the collection

and scoring of writing samples that were used in the ETS study and in the present study. The present study differed from the ETS study in one important aspect that is described in detail in the section which reports the procedures used in the collection of writing samples.

IMPLEMENTATION OF WTR IN FORT WORTH SCHOOLS (FWISD)

Beginning in the fall of 1985, the WTR program was implemented for the first time in 16 elementary schools. In nine of these schools, the program was implemented only with kindergartners. In five of the schools, both kindergartners and first-graders were included in the program. At one school, multi-age classrooms (four- and five year-olds together and first- and second-graders together) were included, while at another school, the program included only first- and second-graders.

Description of the WTR Schools. Of the nine schools using the program with kindergartners only, all but one were minority schools with at least half the children on the free and reduced meal plan. None of these schools had more than four percent of its students who were limited English proficient.

Of the remaining seven schools that used the program with first-graders as well as with other grades, the profile of the schools was much more varied. The percent of students on the free and reduced meal program ranged from 21 to 81 percent, and the percent of students classified as limited English proficient ranged from zero to 76 percent.

Cost of the Program. The total cost of the WTR program for the first year of implementation in 16 schools was $440,508, for a cost per student of $236. The cost covered the following items: salaries for a full-time aide in each of the 16 schools, 8 IBM PC jrs. per school, WTR program software, teacher manuals, student work journals, tape recorders, and other materials and equipment required as part

of the program.

The cost of the program for subsequent years is estimated to be about half the cost of the first year, to cover the recurring costs of the aides' salaries, the student work journals, and maintenance, repair, and replacement of equipment.

IMPLEMENTATION OF WRITING PROCESS IN FWISD

Unfortunately, a very limited number of teachers at the kindergarten and first-grade level had been trained in WP in FWISD at the time of this study. In fact, only one kindergarten and one first-grade teacher were identified. In order to gather more data for this comparison, the evaluation team decided to include two first-grade classrooms in Arlington I.S.D. whose teachers had previously received WP training.

Description of the WP Classes. The kindergarten WP classroom was in a school in which more than half the students were on the free/reduced meal plan and over 20 percent of the students were limited English proficient. The first-grade WP class in FWISD was in a school of the same profile as that of the kindergarten class. The two first-grade WP classes in Arlington I.S.D. were in schools with relatively few children who were on the free/reduced meal plan or who were limited English proficient.

Cost. The only expense incurred in implementing WP in a classroom was the training of the teacher. A 30-hour training session was the minimum required, estimated to cost approximately $32 or less per teacher.

THE FWISD WRITING ASSESSMENT STUDY

Design of the Study. Because of the limitation in the number of available early childhood WP classrooms, it was necessary to match WTR and traditional classrooms to those WP classrooms. Conducting a study with matched groups has a number of

drawbacks. For one thing, due to the lack of pre-treatment measures related to the program objectives, the selection of characteristics on which to match is necessarily questionable. Even in the most idealistically perfect match between groups on a characteristic such as socioeconomic status, the researcher cannot be certain that the groups were matched on a characteristic that equated the groups with respect to the outcome measure. And in the present study, the matches on the selected characteristics were far from perfect. When comparisons are between groups that were different or unequal before the treatment, and when post-treatment differences in the dependent variable are found, one cannot determine whether these differences are a result of the treatment or due to the prior inequalities. As Cronbach (1963) points out:

> Any failure to equate the classes taking the competing courses will jeopardize the interpretation of an experiment and such failures are almost inevitable. (p.49)

Despite the problems caused by the use of non-randomly selected comparison groups, the evaluation team felt that the collection of any information relative to this comparison, no matter how limited, was preferable to no such comparison.

In addition to the one kindergarten and three first-grade WP classrooms each being matched to a WTR and a traditional classroom, additional WTR and traditional classrooms were selected to represent the distribution of WTR classrooms across the district. A total of 24 classrooms were selected to include in the writing assessment study: 10 WTR, 10 traditional, and 4 WP classrooms. Table 1 presents the design of the study.

The schools were matched as closely as possible on two variables: (1) socioeconomic status, as determined by the percent of students on the free and reduced meal program, and (2) the percent of students classified as limited English proficient (LEP). Once the school sites were identified, individual classrooms were selected according to the following guidelines: (1) no ESL or bilingual

classrooms were included, (2) the ethnic makeup of the class was matched as closely as possible for those classrooms matched to a WP classroom, and (3) the classrooms were randomly selected for the additional (unmatched to WP) comparisons. In order to match the sample group as closely as possible for the WP comparison, a traditional classroom from the same school as the WP classroom was chosen for two of the comparisons.

Because the WP teachers might be considered "self-selected," in that these were teachers who voluntarily chose to receive WP training, there was some concern that the "specialness" of these teachers might affect their students' writing scores more than the WP instruction itself. However, it would have been highly inadvisable to try to select the comparison teachers in any but a random manner. The most important consideration in this decision was the fact that any matching procedure would have reduced the generalizability of the results to only the "best" teachers. In other words, in examining the differences between writing skills of students in WTR and those in traditional classrooms (which is the primary goal of the writing assessment study), one could not attribute any found differences to the use of the WTR Program, but rather, only to the use of the program by the best teachers. This is obviously not a very useful generalization. Only with a random selection of classrooms could one make generalizations about the program across levels of teachers. The limitations of the WP comparison are acknowledged, but to spread these limitations to the other comparison would have been devastating to the study.

Table 2 presents the demographic information for those sites selected for the study. These data are based on figures from the Fall, 1985, FWISD School Profiles, and reflect schoolwide characteristics rather than the characteristics of the particular students in the study.

Collection of Writing Samples. Writing samples were collected from each of the 24 classrooms included in the study, with a total of 215 kindergarten and 270 first-grade samples collected. All samples were collected during the morning of Thursday, May 1, 1986, with the exception of two classrooms whose samples were collected on the morning of May 2, 1986.

The procedure followed for the collection of writing samples was very similar to that used in the ETS evaluation of WTR (Murphy and Appel, 1984) with one important difference. Unlike the ETS study in which the classroom teachers administered the instructions and collected the samples, outside evaluators in this study conducted the actual administration of the instructions and collection of writing samples. This procedure had several advantages over the ETS approach. All the evaluators participated in a 45-minute training session to assure some uniformity in the collection procedure. In addition, the presence of outsiders offered some assurance that the teachers in fact did not assist their students in the writing task. (In one classroom, a teacher was observed assisting two students, whose samples were not included in the study.)

A total of 12 evaluators participated in the collection of writing samples. Each evaluator collected the writing samples in two classrooms. To avoid any systematic differences in the administration of the instructions, all but two of the evaluators were assigned to a classroom in two different treatment groups. For logistical reasons, two of the evaluators collected samples from classrooms in the same treatment groups, but both had had extensive experience with elementary students.

The instructions used to elicit the writing samples (see Appendix A) were developed by one of the co-authors who had worked at ETS at the time of the original evaluation of WTR (Murphy and Appel, 1984). The instructions were very similar to those used in the ETS study. If for any reason the instructions used

to obtain the writing samples were biased against any of the children because of their socioeconomic status or their language background, the effects would have been equally distributed across the treatment groups as a result of the matching design.

The instructions were followed exactly as listed in Appendix A by all evaluators with one exception: the time allowed for completing the writing task. In 10 of the classrooms, a few children took longer than the 30-minute period: children in four classes took up to 35 minutes, children in another five classrooms took up to 47 minutes, and one child took an hour and 15 minutes to complete the task. Of these 10 classrooms, five were WTR classes, two were WP classes, and three were traditional classes. Evaluators stated that they rushed a total of 17 students to finish their stories, an average of 1.8 from nine different classrooms. The approximate average amount of time that students from different treatment groups took to complete their stories is listed in Table 3.

The evaluators were asked to describe the students' responses to the writing assignment. Overall, the response was as expected, with WTR and WP students excited and comfortable with the task, while students in traditional classes were somewhat less excited and more confused by the task. There were some notable exceptions to this trend, however. In two of the WTR kindergarten classes, the students acted as if they did not know what to do, and in three of the traditional first-grade classes, the students were quite excited and obviously experienced in creative writing tasks. In addition, the teacher's response in one traditional kindergarten class was fascinating: she said that she had never asked her students to write stories and was shocked to see that they could. When the evaluator asked the students to read their stories to her, the teacher thanked the evaluator because the teacher felt she herself would not have been able to decode their phonetic spelling.

The role of the classroom teacher during the writing assignment varied somewhat. All the teachers were asked to remain in the classroom during this time and to write the lead sentence on the board. Beyond that, the teachers chose different levels of involvement within our guidelines: some just did work at their desks, some helped pass out writing paper, some answered questions, some walked around the room offering encouragement to the students. There was no discernable trend in the teachers' behavior according to treatment group or age of student. Of the 24 classrooms, three had substitute teachers that day: one first-grade traditional and two first-grade WTR classrooms.

The role of the evaluator was quite specific and is outlined in Appendix A. In addition to these written guidelines, all the evaluators were instructed to walk around the room during the writing assignment and offer appropriate encouragement. Most of the evaluators accepted completed papers before the end of the 30-minute session; two evaluators left the papers with all the students for the full 30 minutes.

Scoring of the Writing Samples. The scoring procedure utilized in this study was the same as that used in the ETS evaluation of WTR (Murphy and Appel, 1984). As described earlier, this approach is somewhat holistic in nature, but allows for the decoding of individual words which are spelled phonetically. Overall, the scoring criteria do not include penmanship, spelling, grammar, or punctuation. Rather, the criteria address aspects such as the development and organization of ideas, and other standards reasonable to apply to the writing of young children.

A team of four people (a writing consultant who worked at ETS at the time of the WTR evaluation, a FWISD writing expert, and two FWISD evaluators with extensive early childhood experience) reviewed the writing samples to establish the criteria for scoring, without knowledge of each sample's origin. The kindergarten samples were examined separately from the first-grade samples. The

review team read through all the kindergarten samples and a random group of over half the first-grade samples to establish the criteria. The process was the same for both sets. Upon reading a sample, each reviewer would decide at which level that paper might fall on a scale from zero to six. As each team member read more papers, he or she would begin to formulate a concept of what aspects of a paper would determine the level of that paper. All team members discussed their ideas throughout the process and slowly refined the criteria with team consensus. The final kindergarten scoring criteria are listed in Appendix B, and the final first-grade scoring criteria are listed in Appendix C.

Once the scoring criteria were established, the review team selected a representative writing sample for each scoring level within each grade level to serve as an illustration of that level. The example kindergarten writing samples are included in Appendix D, and the example first-grade writing samples are in Appendix E.

Kindergarten and first-grade teachers from across the school district were solicited to be scorers of the writing samples. A letter requesting teachers' participation was sent to every kindergarten and first-grade teacher of the 45 non-Writing to Read schools. (No teacher from a school included in the study was to be a scorer. Due to timing limitations, the evaluation team did not know which WTR schools were to be included in the study at the time the letter of request for scorers was sent.) A total of 111 teachers responded to the letter. Of these, 26 were from schools included in the study and could not be scorers. Of the remaining 85, 50 were selected to participate. They were selected so as to represent as many different schools as possible and then randomly within this guideline. Of these 50, 38 actually participated in the scoring procedure, representing 25 different schools. Of the 38 scorers, 21 were kindergarten teachers and 17 were first-grade teachers. Because of the larger number of

first-grade samples, six of the kindergarten teachers scored first-grade samples while the rest scored samples of the grade level they taught.

None of the scorers knew the exact purpose of the activity, other than as a study of the development of writing skills of young children. After an introduction, the scorers of kindergarten samples were separated from the first-grade scorers. Each group was then trained on the scoring procedure. Each scorer was given a copy of the set of samples selected to represent the seven different levels for that grade level (see Appendices D and E). The samples were coded so that the scores were unknown, and placed in the following order in each packet: the "6" sample, the "1" sample, the "5" sample, the "2" sample, the "4" sample, the "3" sample, and the "0" sample. The scorers were instructed to read each sample in turn and to place them in piles of "high" and "low" quality. Following this, the scorers were asked to rate each paper with a score from zero to six, in a comparative fashion.

When the group had completed this activity, a hand count was taken to determine how the scorers rated each sample. Appendix F presents this tally of the training samples for the kindergarten scorers, and Appendix G presents this tally for the first-grade scorers. In only one case did any scorer deviate more than two points in either direction from the intended score. The level of consensus for the intended scores ranged from 62 to 94 percent for the kindergarten samples, and from 58 to 100 percent on the first-grade samples. A group discussion of the merits and limitations of each paper followed this tally.

Upon completion of this exercise, each scorer was given a folder with writing samples -- each of the 15 kindergarten scorers received a folder with 14 or 15 random writing samples, and each of the 23 first-grade scorers received a folder with 11 or 12 random samples. Each sample was coded in such a way that the treatment group and school were unknown to the scorers. The scorers were

instructed to use the guidelines outlined for that grade level for scoring (see Appendices B and C) as well as the sample papers. They were told to ignore penmanship, grammar, punctuation, and spelling, and to not compare papers within the folder but rather to compare each to the guidelines and samples. For each of the groups, two members of the original review team were available to answer questions and help scorers to decode words that were spelled phonetically.

Each paper was read and scored independently by two readers. Whenever the two scores differed by three or more points, (e.g., 6-3, 5-2, 4-1), the paper was coded as "discrepant" and read by a third reader. Of the 215 kindergarten samples, 10 (or less than 5 percent) were scored discrepantly, and of the 270 first-grade papers, 20 (or 7 percent) were scored discrepantly. In both cases, this is rather high inter-rater reliability.

The final score for each sample was the sum of the scores given by the two readers. In the case of a discrepant paper that was read by a third reader, the two closest scores were used to determine the sample's final score. Thus, the range of scores for each sample was from zero to 12.

Results of the Writing Assessment. The results of the writing assessment were examined in three ways at each grade level: (1) comparing scores of students in matched WTR, WP, and traditional classrooms, (2) comparing scores of students in WTR and traditional classrooms that were not matched to Writing Process classes, and (3) comparing scores of all WTR and traditional students in the study.

The mean scores on the writing assessment for the students in the matched comparison of WTR, WP, and traditional classes are presented in Table 4. Cautions concerning this comparison must be emphasized. The matching procedure alone is problematic, as discussed in the section on the design of the study. In addition to the statistical and design problems built into this procedure, the range of demographic characteristics of the WTR schools was so limiting that some

of the matches were questionable (see Table 2). A further concern is that even though the level of analysis is at the level of the student, in fact only one kindergarten class and three first-grade classes were included in each of the appropriate cells. Given the important effect that the teacher has on outcome measures such as that used here, this is a considerable limitation. In addition, two of the first-grade WP classes were in another district, and the teacher in the third WP first-grade class was out on maternity leave for eight weeks during the school year.

Because of all the limitations on these data, no tests of significance were conducted. At the kindergarten level, the trend in the data indicated that WTR students scored the highest, WP students next highest, and the students from the traditional class scored the lowest on the writing assessment. Not much weight should be placed on these findings, however, because the comparison was based on only one classroom for each treatment, and the demographic match between the WTR class and the WP and traditional classes was poor. At the first-grade level, the students from the WTR and traditional classrooms scored about the same on the writing assessment while the WP students scored considerably higher.

Tables 5 and 6 present these data in a different format: the range and frequency of scores received by each group. In Table 5, which presents the kindergarten scores for the matched groups, the most obvious feature is the small range of scores (from zero to two) in the traditional classroom. The main difference in scores between the WTR and WP classrooms is the much larger number of samples scored "zero" in the WP than in the WTR classroom. It must be reiterated that while the kindergarten WP and traditional classrooms were from the same school and therefore well matched on demographics, the match between the kindergarten WTR and WP classroom was probably the most disparate (see Table 2), with the WTR class from a much higher socioeconomic status school than the WP and

traditional classes. Therefore, it was difficult to draw any generalizations from these limited data.

Table 6, which presents the first-grade scores for the matched groups, is based on more data than at the kindergarten level. In examining the distribution of scores for the three groups, one sees that the WTR and traditional classroom distributions were essentially the same, with the majority of the scores falling in the low to middle range (from 2 to 6). The bulk of the WP distribution, on the other hand, was definitely in the upper range of scores (from 7 to 12). One must remember, however, that two of the three WP classrooms were located in a neighboring district and may not be comparable to our schools. These schools served high socioeconomic students, and they were matched to the FWISD schools that served the highest socioeconomic students available within the treatment groups. Nevertheless, as with the kindergarten matched data, it was difficult to make any definitive statements about the first-grade matched comparisons.

The comparisons between the scores of students in WTR and in traditional classrooms not matched to WP carried fewer limitations and were more clearcut in their results. Table 7 presents the mean scores for the groups in these comparisons. Tests of significance that were applied to each comparison indicated no significant difference between the writing samples of students in WTR and those in traditional classrooms at both the kindergarten and first-grade level.

Tables 8 and 9 present these data in terms of the range and frequency of scores for each group. At the kindergarten level (Table 8), the most interesting difference between the groups was in the number of samples given a score of zero: 26 percent for the WTR students compared to 45 percent for the non-WTR students. On the other hand, 18 percent of the WTR students received a score of four or above compared to 25 percent of the non-WTR students.

In the same comparison at the first-grade level, as depicted in Table 9, very few differences were obvious. Twenty percent of the WTR students received a score of zero on the writing assessment compared to five percent of the non-WTR students. Other than that, the distributions were very similar.

In the third overall comparison, the scores for all the WTR students in the study were compared to the scores for all the students from traditional classrooms in the study. In other words, this comparison combined the students from the first two overall comparisons, but eliminated the students from the WP classrooms in the comparison. Table 10 presents the means and standard deviations for this comparison.

A test of significance at the kindergarten level indicated a significant difference between the scores of WTR students and those in traditional classrooms. This is not surprising when one realizes that the matched WTR score was the highest of all and its matched traditional score was the second lowest of all classes. These two outlier scores changed the overall complexion of the comparison considerably, but because the same limitations of non-random assignment apply, this comparison must be viewed with great caution. At the first-grade level, the difference between WTR and traditional students remained non-significant.

As the standard deviations indicated (see Table 10), there was a great deal of variance among the various classrooms included in the study, particularly at the kindergarten level. Table 11 presents the mean scores for the individual kindergarten sites, and Table 12 presents the mean scores for the individual first-grade sites. As one can see, the variance within each group was at least as large as the variance between groups, particularly at the kindergarten level. These data reinforce the cautions with which one should view the writing assessment results.

Discussion of the Writing Assessment Results. Overall, when reviewing the results of the writing assessment study, one must remember the strong limitations of the study design: a small sample size and non-randomly selected comparison groups. However, within these constraints, the results are straightforward with regard to one comparison and suggest trends on the other comparisons. The one straightforward result which was supported throughout the analyses is that there was no difference between the writing samples produced by the first-grade WTR students and those produced by the first-grade students in the traditional classes. Even within the constraints of the matched group design, assuming that there were inequalities between the groups prior to the treatment, one would strongly expect that the presence of a treatment such as WTR compared to no such treatment would produce some effect on a direct outcome measure. No such effect was even suggested in the first-grade data. Based upon these results, in conjunction with the observations of the evaluators who collected the writing samples, one might conclude that traditional first-grade teachers may in fact be including creative writing as part of their curriculum, at least within the confines of the type of first-grade classrooms included in this study.

The data from the Writing Process classrooms at the first-grade level strongly suggested an effect from this type of instruction, compared to that used in WTR and traditional first-grade classrooms. However, the effect was strongest from the schools serving the highest socioeconomic status students, so that any generalizations were limited. Nevertheless, the data indicated that this cost-effective alternative to improving the writing skills of young children offers a great deal of promise and should be examined in greater depth.

At the kindergarten level, the only comparisons which supported the positive effect of WTR on the development of writing skills were those which included the matched classrooms in the comparison. The WTR school in this match served much

higher socioeconomic status students than the school matched to it, and the resulting scores placed the students from these two schools at the two ends of the writing assessment continuum.

When these poorly matched outlier classrooms were not included in the data, the results indicated no significant differences between the writing skills of kindergartners in the WTR program and those in traditional classrooms. However, with the wide variance within groups as well as between groups, one might conclude that the individual teacher had at least as much effect on the writing skills of his/her kindergarten students as the presence of the WTR program.

In the comparison of WP with WTR and traditional writing instruction at the kindergarten level, again only a trend was suggested. This comparison was limited to only one classroom per group and suffered from the poor match between the WTR school and the school in which both the WP and traditional classrooms were located. However, the match between the WP and traditional classrooms, being from the same school, was as good as it could be on the population demographics. This comparison suggested a positive effect when the teacher had been trained in WP, but again, more extensive investigation is necessary before any definitive conclusion is possible.

It should be noted that the findings of this study differed from those of the ETS study of WTR (Murphy and Appel, 1984), in which differences in writing ability between WTR students and students in traditional classrooms were found. The only difference between these two studies in the implementation of the writing assessment study was in the data collection procedure: in the ETS study, the teachers themselves collected the samples, whereas in the present study, outside evaluators collected the samples. The latter procedure insured the validity of the samples.

CONCLUSION

It must be recognized that it is somewhat difficult to plan, execute, and analyze the results of a well-designed writing assessment study. Nevertheless, the information obtained from such an endeavor is invaluable to evaluators of programs which purport to improve the writing abilities of the participants. We hope that the procedures outlined in this paper will both encourage and guide other school district personnel to utilize direct assessment procedures for evaluating the development of writing skills in their students.

We would like to suggest a few general guidelines in conducting such an endeavor. One of these is to use a team approach in designing a writing assessment study. In our study, combining one team member's expertise in teaching and assessing writing with another's expertise in the creation of a workable and valid research design resulted in a far more informed approach toward evaluating writing than either member could have achieved individually.

Second, those who undertake this type of study should be aware that a well-planned writing assessment must be carefully designed far in advance of the actual scoring of the writing samples. The creation of the prompt, its administration, the collection of the papers, the selection of sample training papers, and the definition and refinement of the scoring criteria are not mere mechanical duties to be executed quickly, but rather are significant variables that can affect the validity of the results if not implemented with care. For example, we believe that the administration of the prompt and the collection of the writing samples by outside evaluators rather than by the classroom teachers was an essential element in the obtainment of writing samples that truly reflected the skills of the writers. In addition, when the steps of the writing assessment are applied to the collection of writing samples from young children, extra caution must be used to address the particular idiosyncracies and needs of

this age writer.

When school district personnel design, administer, and analyze their own direct writing assessments, rather than using some machine scored method provided by a "professional" testing agency, all those involved benefit from some wonderful by-products. The school administrators are able to inexpensively assess writing with the active participation of the district's teachers themselves. The teachers who conduct the scoring gain a sense of method through which they themselves can assess the work of their students and integrate it into their regular instruction on an ongoing basis. The students of these teachers will benefit from both their teacher's increased insight and skills and from their own increased sense of what is realistically expected and possible in successful writing. Such an awareness is fostered at a simple level when their teachers share with them sample training papers which exhibit the characteristics of effective writing by other students of their grade level. At a more complex level, the students will have increased awareness on an ongoing basis as their teachers begin to tie the assessment methods to their instruction in writing. In this way, connections between program evaluation and program development are formed.

Indeed, the use of direct assessment methods of evaluating writing can create a sharing and refinement of knowledge among individuals that will contribute toward a more informed and richer experience for our students.

References

Cronbach, L. (1963). Course improvement through evaluation. Teachers College Record, Vol 64.

Emig, J. (1977). Writing as a mode of learning. College Composition and Communication, 28, 122-128.

Godshalk, F., Swineford, F., and Coffman, W. (1966). The assessment of writing ability. New York: College Entrance Examination Board.

Graves, D.H. (1981). A case study observing the development of primary children's composing, spelling, and motor behaviors during the writing process. Final Report (NIE Grant No. G-78-0174). Durham, New Hampshire: University of New Hampshire. (ED 218 653)

Graves, D.H. 1983). Writing: Children and teachers at work. Exeter, New Hampshire: Heinemann.

Guttinger, H. (1986, Spring). Writing to Read: It's a gold-medal winner. Educators' report. Atlanta: IBM Educational Systems.

Hathaway, W.E. (1985, March). A critique of the ETS "Evaluation of the Writing to Read Instructional System 1982-1984". Portland, OR: Portland Public School District, Research and Evaluation Department.

Karam, P. (1986, June). Writing to Read Program evaluation report. San Antonio, TX: San Antonio Independent School District, Office of Evaluation.

Murphy, R.T., & Appel, L.R. (1984). Evaluation of the Writing to Read Instructional System: Second year report. Princeton, New Jersey: Educational Testing Service.

Myers, M. (1981). A procedure for writing assessment and holistic scoring. Urbana, IL: National Council of Teachers of English.

Norton, D.P. (1984, July). An evaluative report on the IBM Writing to Read Program as implemented in the Wausau School District during the 1983-84 school year. Wausau, Wisconsin: Wausau School District.

Odell, L. & Cooper, C. (1977). Evaluating writing: Describing, measuring, judging. Urbana, IL: National Council of Teachers of English.

White, E.M. (1985). Teaching and assessing writing. San Francisco: Jossey - Bass.

TABLES

Table 3

Estimated Mean Number of Minutes for Students in Different Treatment Groups to Complete their Stories

	Writing to Read	Writing Process	Traditional
Kindergarten	12.2	15	13.4
First-Grade	24	23.3	17.4

Table 4

Mean Writing Assessment Scores for Students in Classrooms Matched to Writing Process Classrooms

	Writing to Read	Writing Process	Traditional
Kindergarten	6.67 (N=21)	3.21 (N=19)	0.61 (N=18)
First-Grade	4.85 (N=53)	7.89 (N=71)	4.72 (N=64)

Table 1

Number of Classrooms from Each Treatment Group Selected for the Study

	Writing to Read	Writing Process	Traditional
Kindergarten	5	1	5
First-Grade	5	3	5

Table 2

Demographic Description of the Schools Included in the Writing Assessment Study

	Writing to Read			Writing Process			Traditional		
		FRM[a]	LEP[b]		FRM	LEP		FRM	LEP
KINDERGARTEN	Site 1 :	26	12	Site 2[c]:	59	21	Site 3[c]:	59	21
	Site 4 :	67	76				Site 5 :	77	53
	Site 6 :	85	4				Site 7 :	85	5
	Site 8 :	51	0				Site 9 :	51	1
	Site 10:	57	0				Site 11:	63	2
FIRST-GRADE	Site 12:	38	21	Site 13[d]:	57	21	Site 14[d]:	57	21
	Site 15:	24	7	Site 16 :	10	3	Site 17 :	12	2
	Site 18:	21	0	Site 19 :	3	0	Site 20 :	14	0
	Site 21:	60	27				Site 22 :	60	14
	Site 23:	81	44				Site 24 :	75	46

a) the percent of students in that school on the free and reduced meal program (FRM)

b) the percent of students in that school identified as limited English proficient (LEP)

c,d) different classrooms from the same school

Table 6

Distribution of First-Grade Writing Scores for Classrooms Matched to Writing Process Classrooms

	Frequency of Scores by Treatment Group		
Scores	Writing to Read (N=53)	Writing Process (N=71)	Traditional (N=64)
12	X X	0 0 0 0 0 0 0	
11		0 0 0 0 0 0	* *
10	X X X	0 0 0 0 0 0 0 0 0 0 0 0 0	* * * *
9	X	0 0 0 0 0 0 0 0 0 0 0 0	* *
8	X X	0 0 0 0 0 0 0	* * * *
7	X X X	0 0 0 0 0 0 0 0 0 0 0 0	* * *
6	X X X X X X X X X X	0 0 0	* * * * * * * * *
5	X X X X	0 0	* * * * * * * *
4	X X X X X X X X X X X X	0 0 0	* * * * * * * *
3	X X X X X	0 0 0 0	* * * * * * *
2	X X X X X X	0 0	* * * * * * * * * *
1	X X	0 0	* *
0	X X X	0	* * * * *

Table 5

Distribution of Kindergarten Writing Scores for Classrooms Matched to Writing Process Classrooms

Frequency of Scores by Treatment Group

Scores	Writing to Read (N=21)	Writing Process (N=19)	Traditional (N=18)
12	X	0	
11	X		
10	X X X X X	0	
9			
8	X X X X	0	
7	X		
	X X	0 0	
5		0	
4	X X	0	
3	X	0 0	
2	X X	0 0	* * * *
1	X		* * *
0	X	0 0 0 0 0 0 0 0 0	* * * * * * * * * * * *

Table 8

Distribution of Kindergarten Writing Scores for Classrooms Not Matched to Writing Process Classrooms

Scores	Frequency of Scores by Treatment Group	
	Writing to Read (N=81)	Traditional (N=76)
12		
11		
10	X	0
9	X	0 0
8	X X	
7	X X	0
6	X	0 0 0 0 0
5	X X X X X	0 0 0 0 0
4	X X X	0 0 0 0 0
3	X X X X X X X X X X X X X X X X	0 0 0 0 0
2	X X X X X X X X X X X X X	0 0 0 0 0 0 0 0 0 0
1	X X X X X X X X X X X X X X X X	0 0 0 0 0 0 0 0 0 0
0	X X X X X X X X X X X X X X X X X X X X X	0 0 0 0 0 0 0 0 0 0

Table 7

Mean Writing Assessment Scores for Students in Classrooms Not Matched to Writing Process Classrooms

	Writing to Read	Traditional
Kindergarten	2.25 (N=81)	2.00 (N=76)
First-Grade	4.13 (N=40)	4.57 (N=42)

Table 9

Distribution of First-Grade Writing Scores for Classrooms Not Matched to Writing Process Classrooms

	Frequency of Scores by Treatment Group	
Scores	Writing to Read (N=40)	Traditional (N=42)
12	X	0
11	X	
10	X	
9		0
8	X X X	0 0 0 0 0
7	X X X	0 0
6	X X X X X	0 0 0 0 0 0 0
5	X X	0 0 0 0 0
4	X X X X X X X	0 0 0 0 0
3	X X	0 0 0 0
2	X X X X X X	0 0 0 0 0 0 0 0 0
1	X	0 0
0	X X X X X X X X	0 0

Table 10

Mean Writing Assessment Scores and Standard Deviations for All Study Participants in WTR and Traditional Classes

	Writing to Read			Traditional		
	Mean	S. D.	N	Mean	S. D.	N
Kindergarten	3.16	3.15	102	1.73	2.37	94
First-Grade	4.54	3.02	93	4.66	2.82	106

Table 11

Mean Writing Assessment Score for Each Kindergarten Site

Writing to Read		Writing Process		Traditional	
Site 1:	6.67	Site 2:	3.21	Site 3:	0.61
Site 4:	1.74			Site 5:	1.58
Site 6:	1.61			Site 7:	0.33
Site 8:	3.91			Site 9:	4.21
Site 10:	1.67			Site 11:	2.06

Table 12

Mean Writing Assessment Score for Each First-Grade Site

Writing to Read		Writing Process		Traditional	
Site 12:	6.29	Site 13:	5.00	Site 14:	5.00
Site 15:	3.94	Site 16:	8.85	Site 17:	4.53
Site 18:	4.39	Site 19:	9.04	Site 20:	4.61
Site 21:	5.19			Site 22:	5.45
Site 23:	2.95			Site 24:	3.77

APPENDICES

Appendix A

Instructions for Collection of Writing Samples

1. Materials: Students should use the writing implements that they normally use and the Primary Writing Tablet paper that has been provided. If they do not have anything to write with, supply each with a pencil.

2. Length of Time: After the directions have been given, allow the students 30 minutes to complete their story. If they finish early, collect the sample and give those who are finished extra paper for drawing a picture.

3. Length of Response: There is no required length.

4. Teacher Assistance: Please do not assist the students in any way. If the student needs more than one sheet of paper, provide that, but do not staple the pages together until after the 30 minute period.

5. Directions: After the pencils and paper have been distributed, give the students the following directions:

   (Say) Today I'm going to ask you to write a story all by yourself. I want you to tell a story about a make believe ride in a magic car. You can do anything you want to on the ride. You can drive to the moon, or you can drive back in time to see the dinosaurs. You can drive to visit someone that you like or you can go to eat chocolate ice cream with Big Bird. You can pretend to go anywhere that you want.

   Now, I want you to write a story all by yourself about what you would do if you took a ride in a magic car.

   Your teacher is going to write the beginning of the story on the board:

   (Print out this stem on the board):

   Once I took a ride in a magic car.
   (Read this sentence aloud to the children.)

   (Say) Now you write the rest of the story yourself. Try it!

6. Special Circumstances:

   a) If any children ask you to repeat the directions, or say they do not know what to do, repeat the last sentence of the prompt:

      "I want you to write a story all by yourself about what you would do if you took a ride in a magic car."

   If any children ask if they can draw a picture, say:

      That's o.k., but we'd like you to write a story too if you can."

c) If any child becomes visibly upset by the task, you may go to that individual child and say quietly:

"It's o.k. to draw a picture with your story if you want to."

d) If any children ask for help from other students or the teacher, say:

"Try to do it by yourself."

7. Concluding the assessment: At the end of the 30 minute period, take up all papers. If the student has used more than one sheet, staple the sheets together. Staple the index card with the code facing down at the same time. Put the staple at an angle in the top left hand corner. Thank the children and the teacher for their time and cooperation!

Appendix B

Kindergarten Scoring Criteria

Level 6

The writer clearly conveys a narrative account through vivid or striking word choice. The writer may also indicate an awareness of the cause and effect relationship of events.

Level 5

The writer conveys a narrative account through clear word choice.

Level 4

The writer conveys an event or events, but these may be somewhat disjointed.

Level 3

The writer attempts to convey an event or events but often falls into a pattern of repetition or listing.

Level 2

The writer establishes a direct relationship with the stem, but little or no detail is given.

Level 1

There is some attempt to write, but there may only be one intelligible word beyond the stem.

Level 0

The stem is merely copied, or the stem is copied and a picture is drawn. Unintelligible words may follow.

Appendix C

First-Grade Scoring Criteria

Level 6

The writer clearly conveys a narrative account through vivid or striking word choice. Details are essential to the sense of a narrative.

Level 5

The writer conveys a narrative account through clear word choice. Details contribute to the sense of a narrative.

Level 4

The writer has a simplistic form of narrative, although details may be somewhat additive or repetitious.

Level 3

The writer attempts to convey a sense of story, but the story is often repetitious in its detail.

Level 2

Some details are added beyond the stem, but these may be very disjointed.

Level 1

Merely a comment or a list of words follows the stem.

Level 0

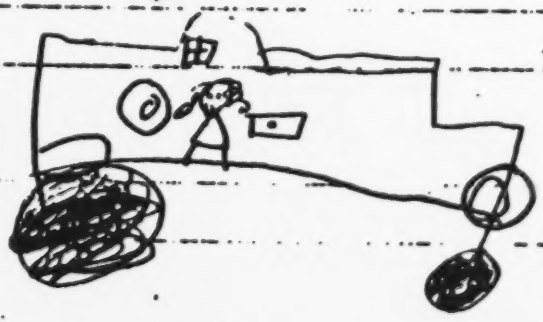
Material following the stem does not constitute an understandable sentence.

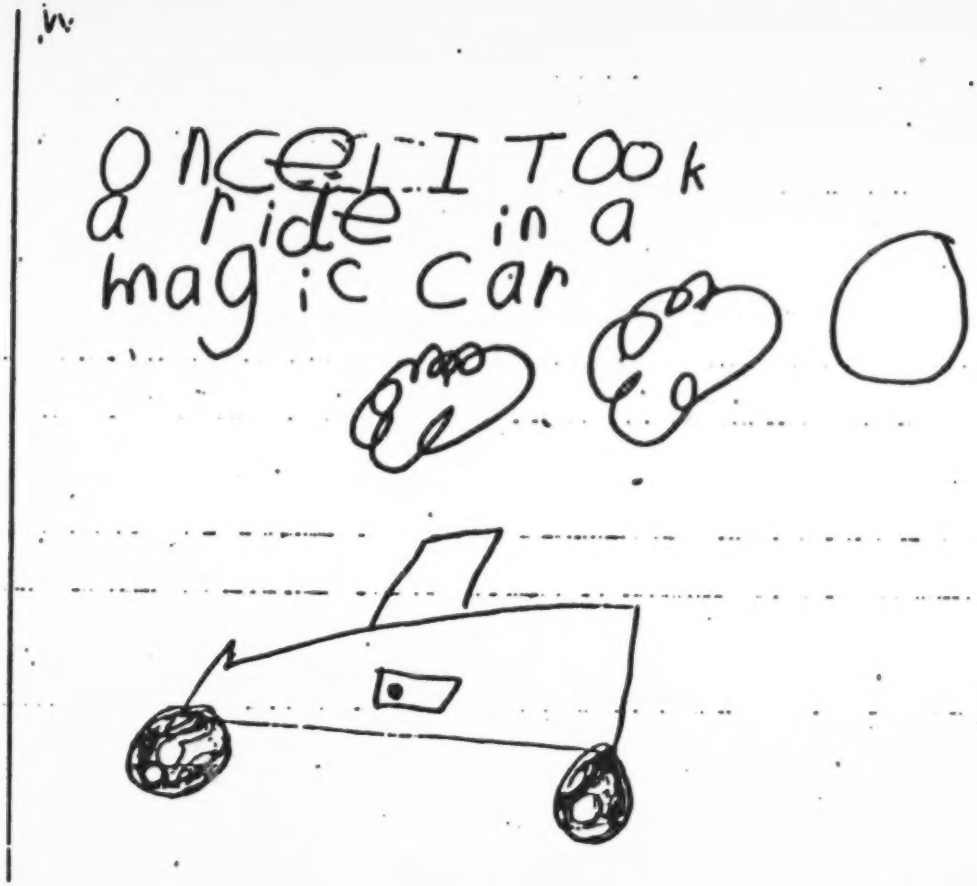
Appendix D

Examples of Kindergarten Writing Samples

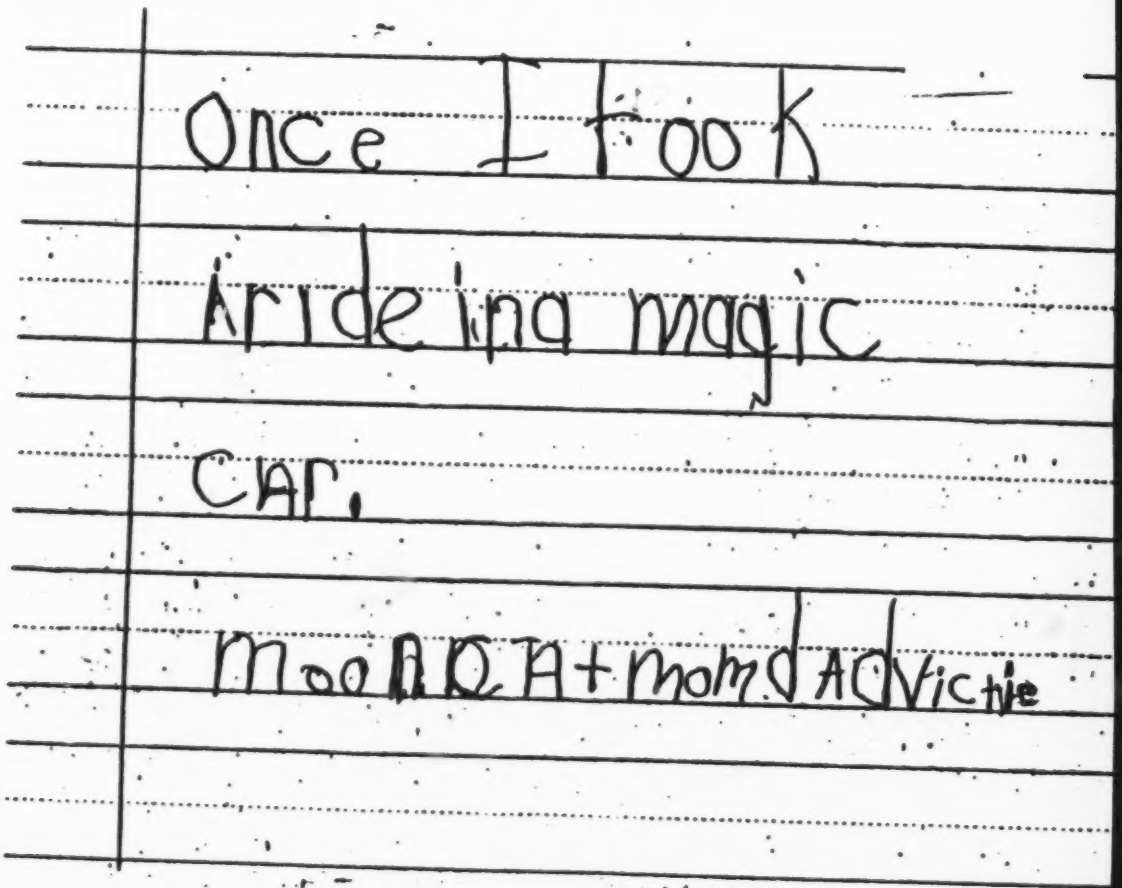
Kindergarten Writing Sample Score of 0

once I took
a ride in a
magic car

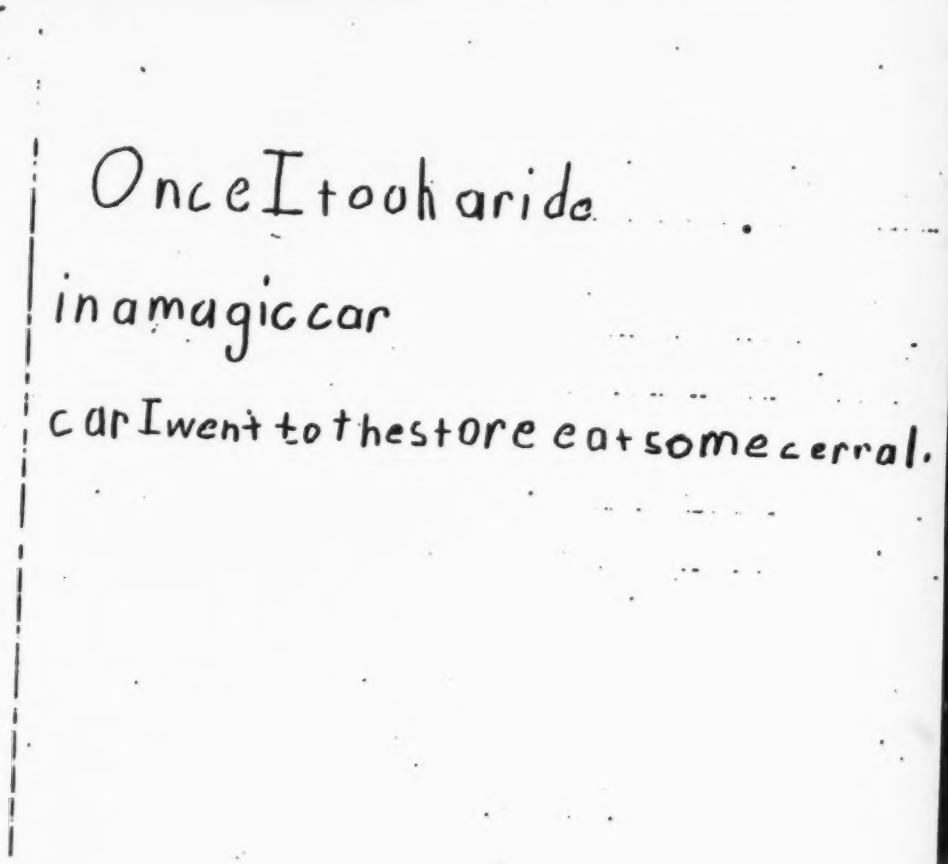



ONCE I TOOK
a ride in a
magic car

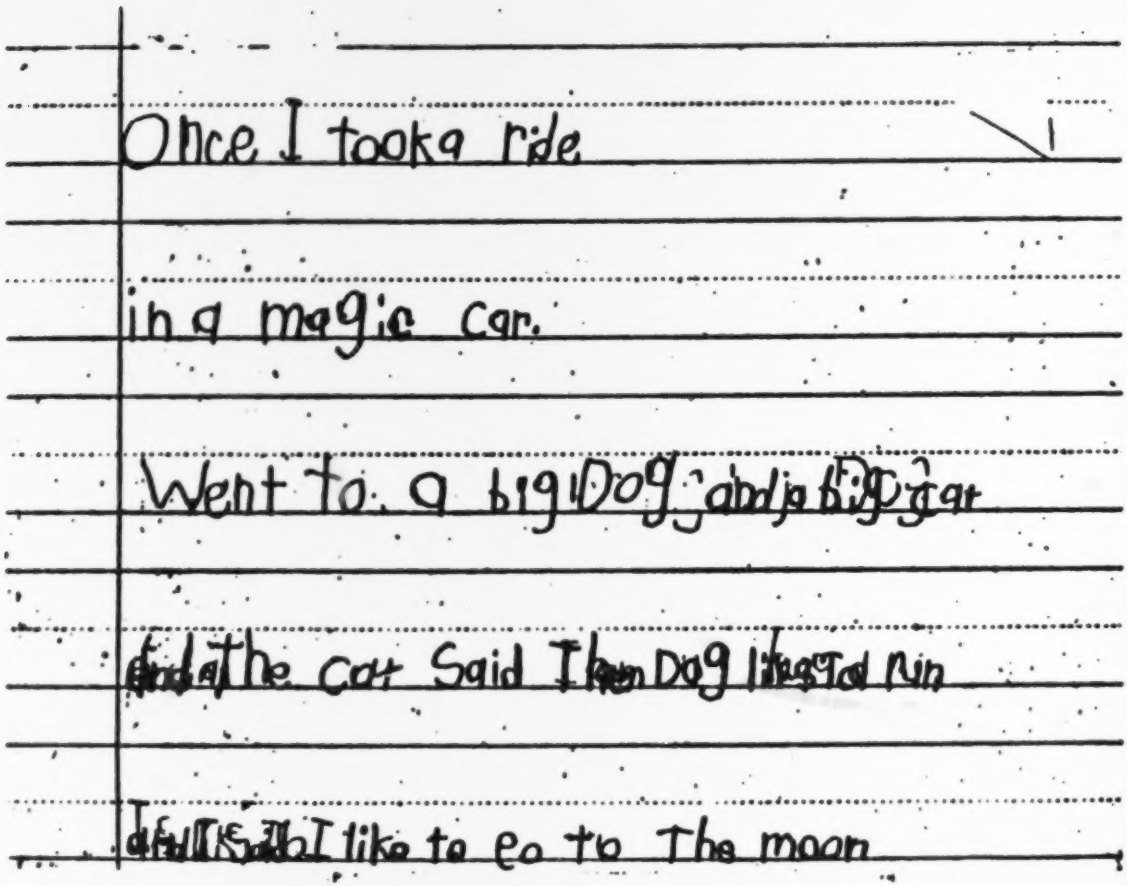
Kindergarten Writing Sample Score of 1



Kindergarten Writing Sample Score of 2



Kindergarten Writing Sample Score of 3



I Said I like you The cat Said I like you to

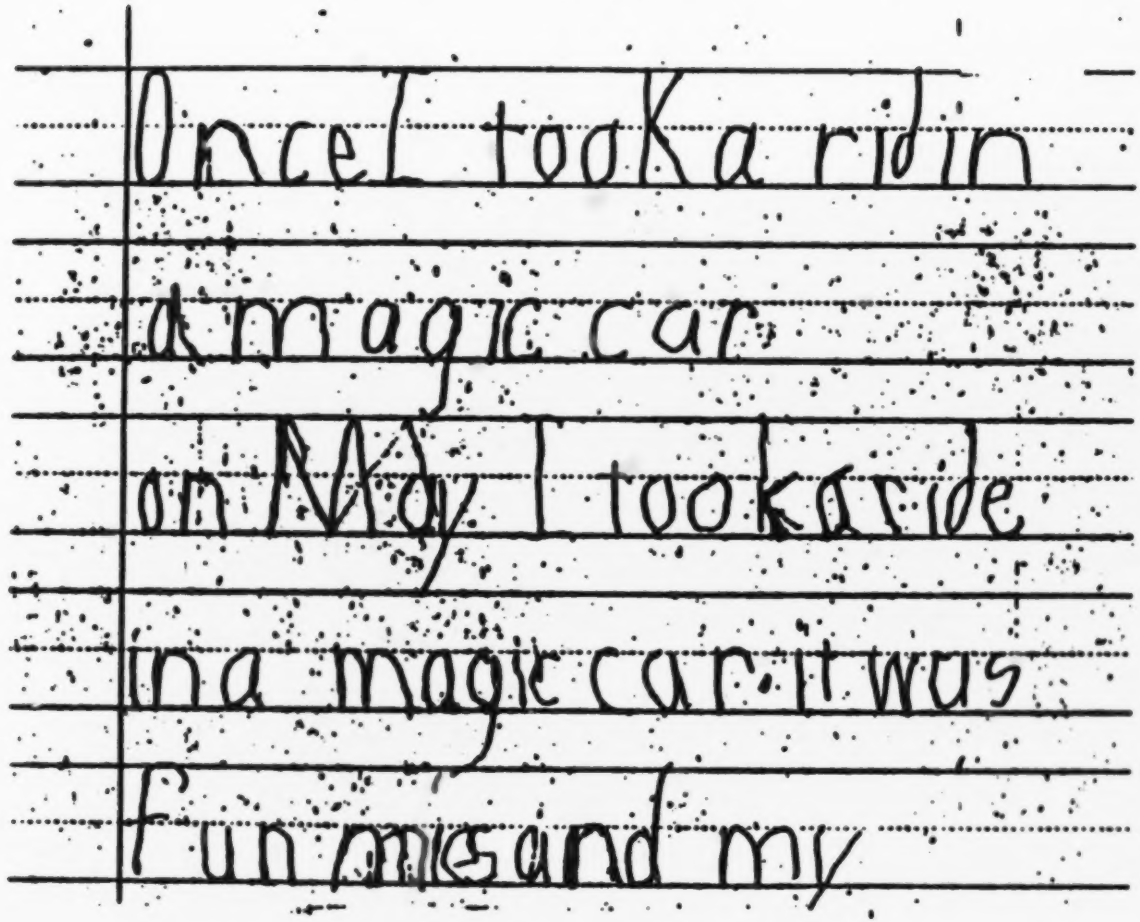
The Dog Said i like you to the Dog

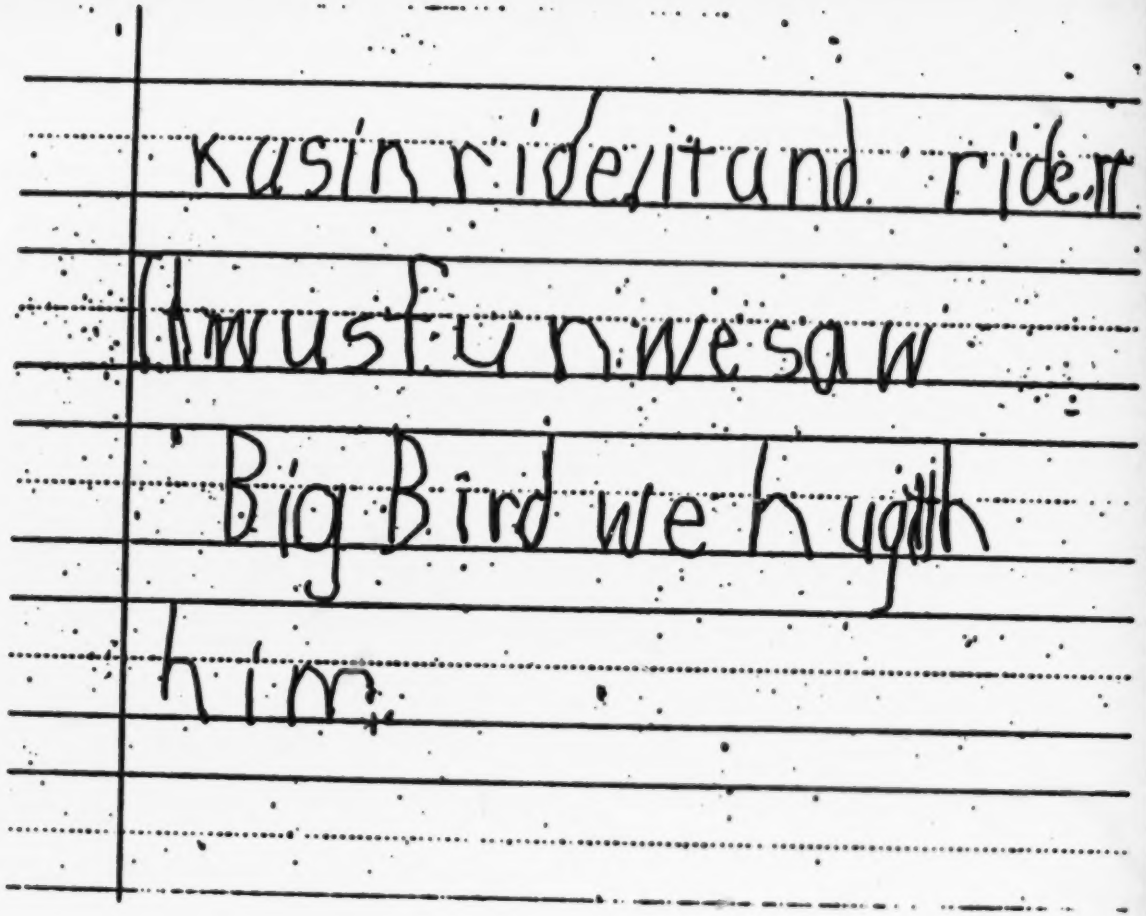
went to The car The car Said

i like you The Dog Said I like

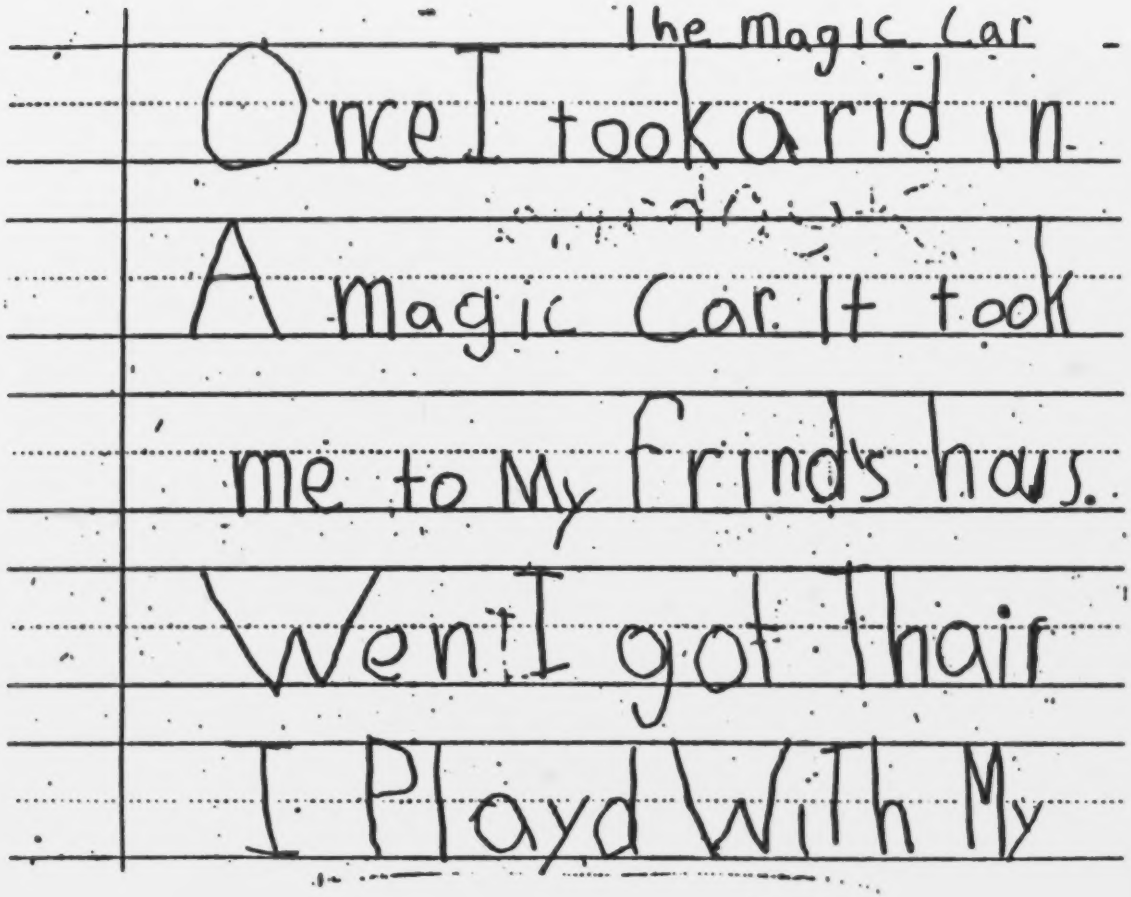
you to The Dog Said I love you.

Kindergarten Writing Sample of 4




kasin ride it and ride it
thwus fun we saw
Big Bird we had with
him.

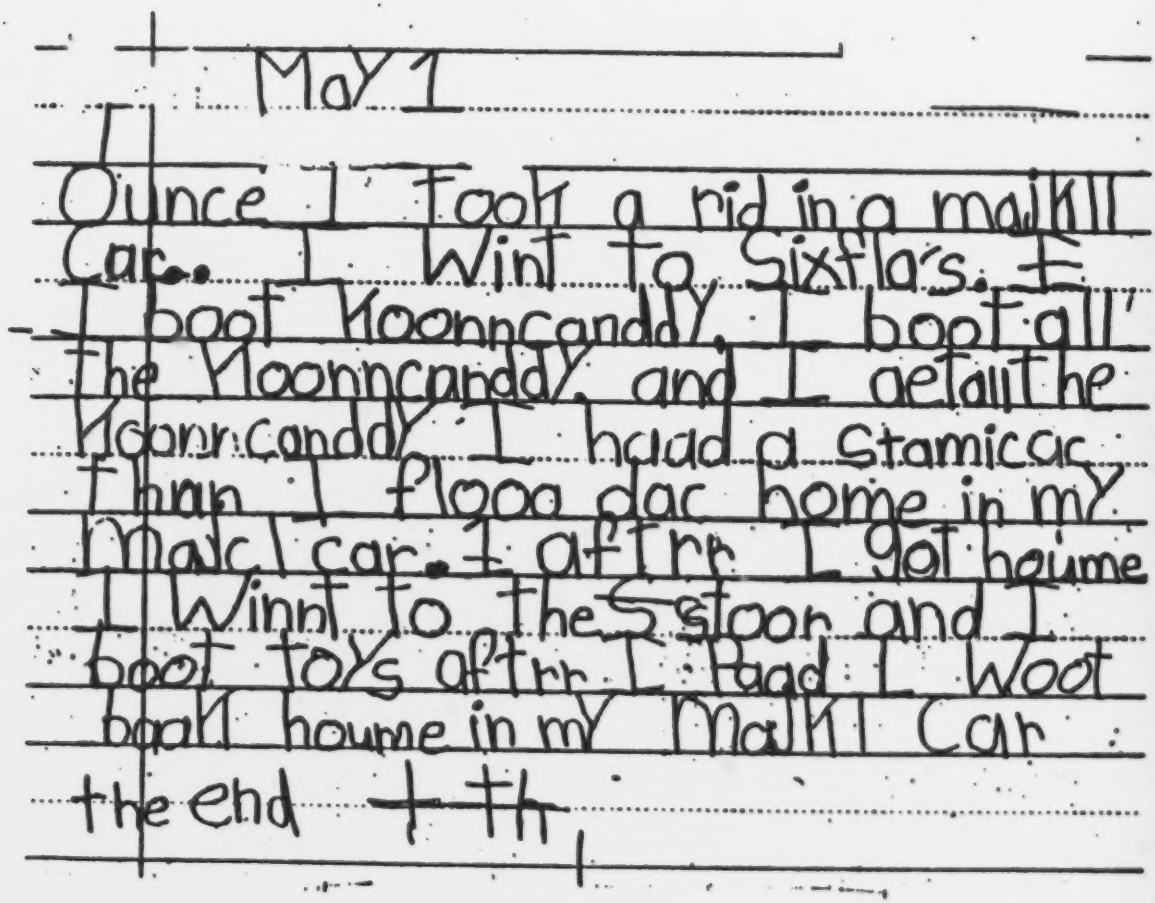
Kindergarten Writing Sample Score of 5



Frind in The Magic Car. The End

May 1, 1986

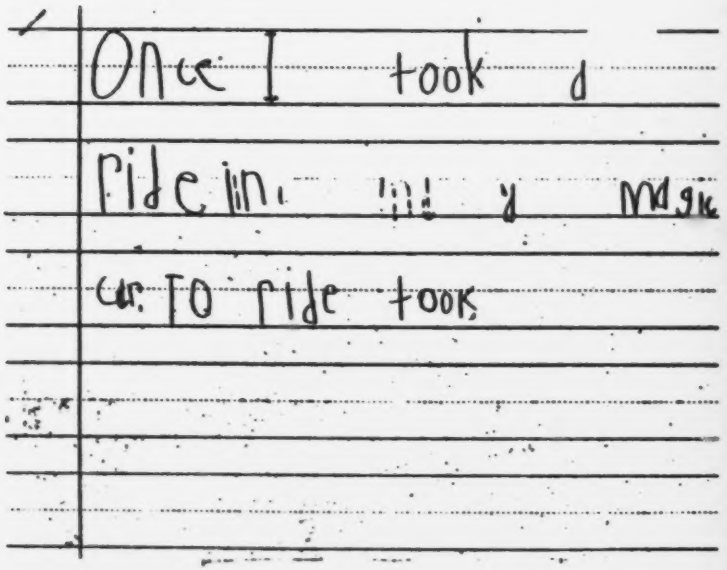
Kindergarten Writing Sample Score of 6


May 1
Ounce I took a rid in a maikit
Car. I Wint to Sixflars. E
I boot Hoonncandy. I boot all
the Hoonncandy. and I aetall the
Hoonncandy I haad a Stamicar
thran I flooa dar home in my
Maic car. + aftrr I got haume
I Winnt to the Stoon and I
boot toys aftrr I Paad I Woot
boak houme in my Maikit Car
the end

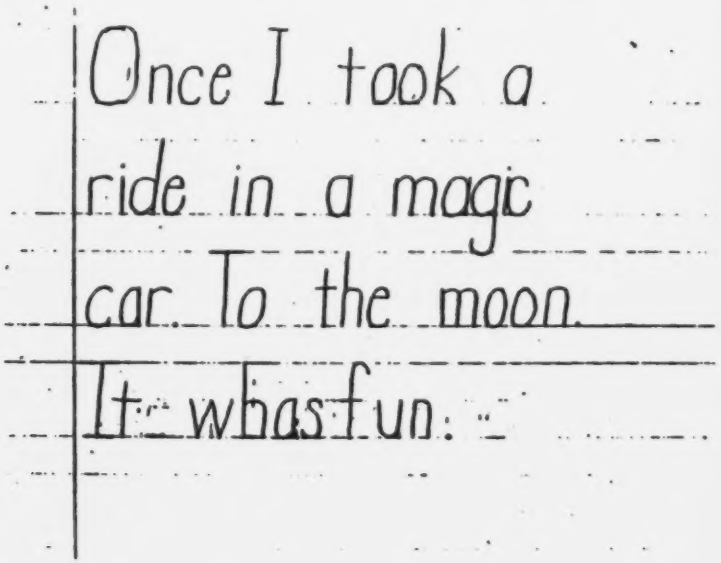
Appendix E

Examples of First-Grade
Writing Samples

First-Grade Writing Sample Score of 0

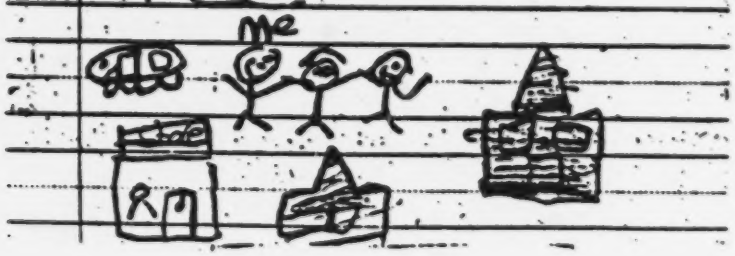


First-Grade Writing Sample Score of 1



First-Grade Writing Sample Score of 2

Once I took a ride in a magic car. I went to see my friends. my friend live in a other state. mexico is a small tree.



First-Grade Writing Sample Score of 3 Page 1 of 2

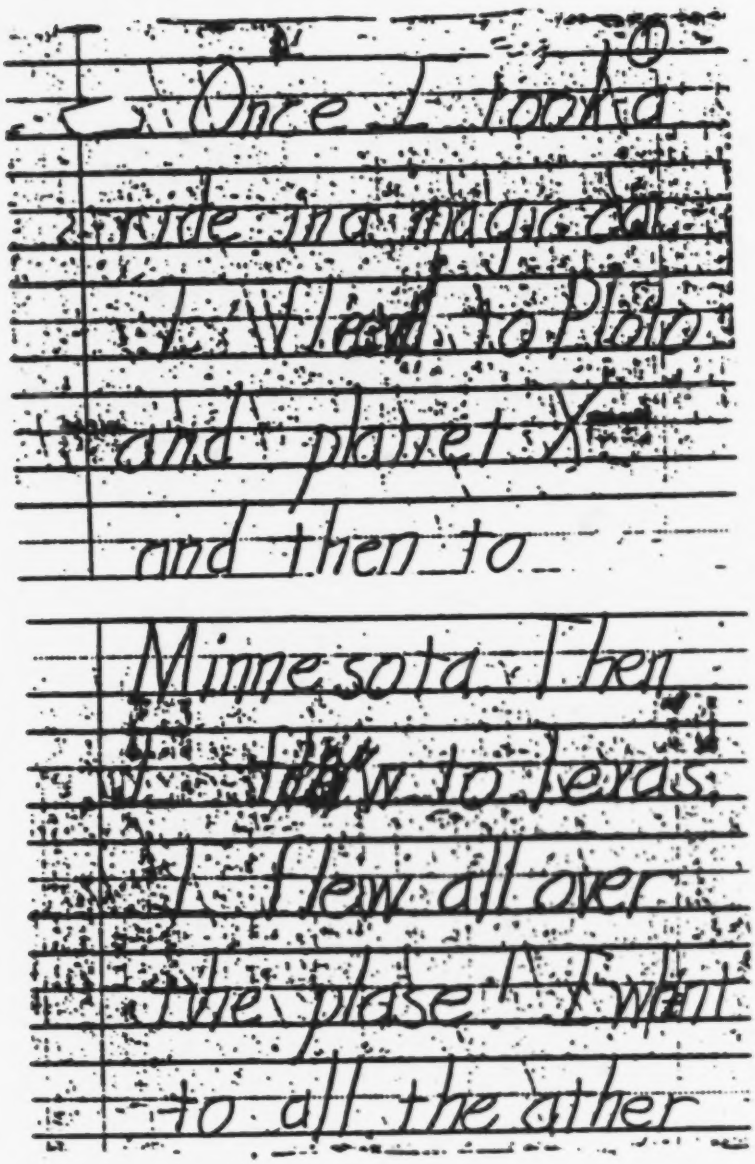
May.1,1986

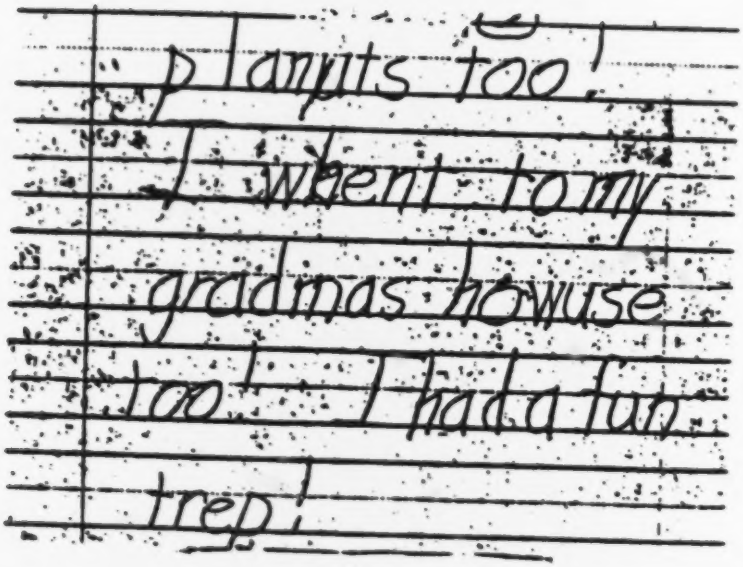
Once I took a ride
in a magic car and
I waent to BigBirds
house and we waent
to the moon and
we waent to see my
mom and we waet
to see my dad to and
we waent to see my

sister and we wantto
see my brother and
we waent to see
Big Birds mom and
dad and siser and

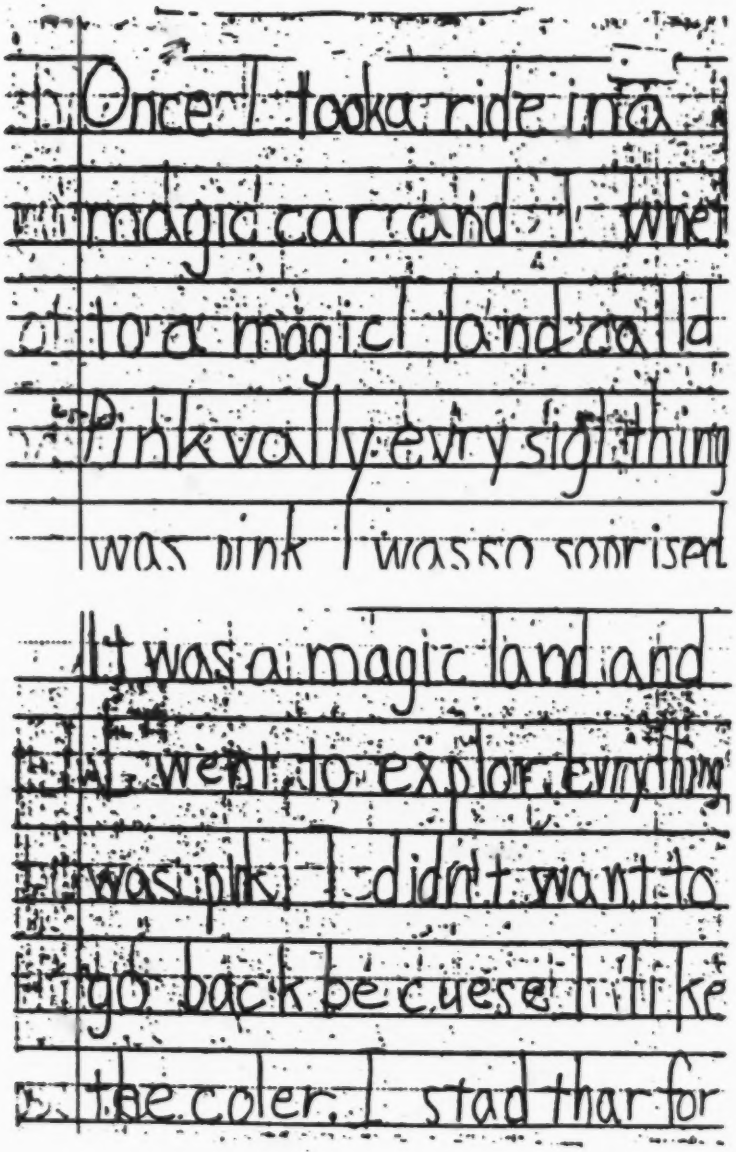
brother and we
waent to get some
icecrem and we
had fun.
The. End.

First-Grade Writing Sample Score of 4 Page 1 of 2




plants too!
I went to my
gradmas howuse
too. I had a fun
trep!

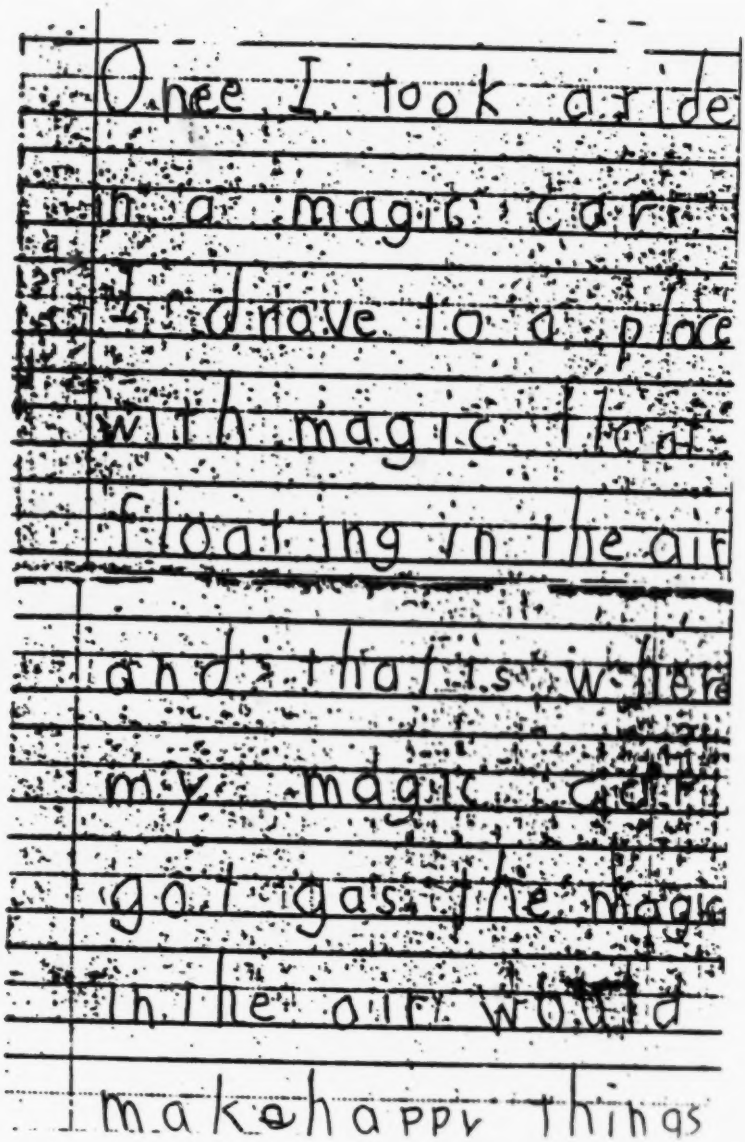
First-Grade Writing Sample Score of 5 Page 1 of 2


Once I took a ride into a
magic car and I whet
to a magic land and it
Pink vally evry sigt thing
was pink I was so sorrised
It was a magic land and
I went to explor evrything
was pik I didn't want to
go back because like
the coler. I stad thar for

meny days and evrything
and evry where was pink
I had so much fun but
I wanted to go back
home so I packed up

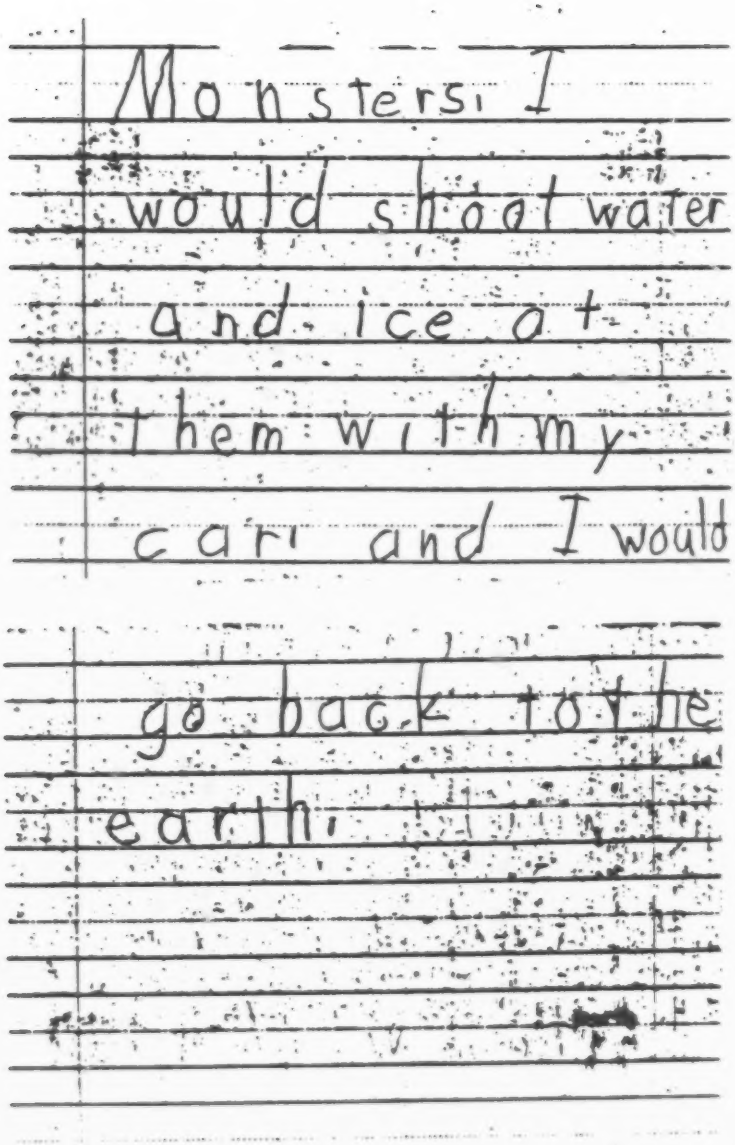
turth was three
planits from here I
was there by 9:00 o'clok
and I was glad to be
home THE End

First-Grade Writing Sample Score of 6 Page 1 of 4



happen. My car
would zoom into
timewarp in till
I went to the
futare, And I wald
travel to china
and see a dragon
and see men sowrd
fihgt I would
travel to Neptune

and see aleyens
wit black panth
heads and lion
bodys. they
would trie to
eat me and my
magic Car would
shoot them and
I'd ride to the
sun and see fire


Monsters. I
would shoot water
and ice at
them with my
car and I would
go back to the
earth.

Appendix F

Tally of Scores on Kindergarten Training Samples

Intended Scores	Scores Given by Trainees*						
	6	5	4	3	2	1	0
6	14	2					
1					3	13	
5	1	10	5				
2				4	12		
4		1	10	4	1		
3	1	1		14			
0						1	15

*The total number of kindergarten trainees included one central office administrator who did not participate as a scorer.

Appendix G

Tally of Scores on First-Grade Training Samples

Intended Scores	Scores Given by Trainees*							
	6	5	4	3	2	1	0
6	18	4	2				
1				1	6	17	
5	7	14	2	1			
2			1	2	16	5	
4		5	15	4			
3		1	3	17	1	2	
0							24

*The total number of first-grade trainees included one central office administrator who did not participate as a scorer.